



Foreword

Database Theory 2004

This special issue of JCSS is dedicated to selected papers in database theory over the last two years. We have tried to give a range of papers covering most of the major themes within recent database research. Still, there is one overarching “meta-theme”: the vast impact of the networking revolution on data-processing. Most of the research presented here is driven by the need either to make data available over networks or to process the large volumes of data that have become available over the network.

Networked databases are dealt with most explicitly in the paper by Bawa, Gionis, Garcia-Molina, and Motwani and by Cohen and Kaplan. Bawa et al. consider algorithms for processing queries aggregating data over highly dynamic networks—networks where peers can come and go during the course of query-processing. Cohen and Kaplan’s work also confirms that query-processing in a network setting cannot be exclusively defined in terms of traditional mappings between database instances. Their paper analyzes a particular kind of “continuous query”—time-decaying aggregates—that arises in monitoring network streams.

Probably the largest impact of the networked world on database systems has been in the area of data exchange. The eXtensible Markup Language (XML) has emerged as the predominant data format for data exchange, and much recent research work has dealt with mechanisms for defining and evaluating queries, transformations, and views taking data into and out of XML formats. The survey by Thomas Schwentick gives an overview of one of the most important components in the foundations of XML processing—tree automata. It describes the role that tree automata play in XML schemas, XML path queries, and XML transformation. The paper by Davidson, Fan, and Hara and that by Martens and Neven deal with mappings acting on XML documents. Davidson, Fan, and Hara are concerned with storage of XML in relations. They study the ability to generate keys in relational databases populated from an XML document through a storage mapping. Martens and Neven investigate XML-to-XML transformations, focusing on statically analyzing type-correctness of transforms given in a restricted subset of XSLT. The paper by Bar-Yossef, Fontoura, and Josifovski is related both to new query-processing models and to data exchange. Their work examines the processing of XML queries in the XPath language over streams. Stream-based evaluation of XML queries is applicable in the setting where XML data flows in from a network and when XML-processing is done on static documents on top of a SAX parser.

Support for data manipulation in a networked setting is also needed at the application level. The paper by Deutsch, Siu, and Vianu proposes high-level language support for describing web-based services that interact with data. Their work provides both a specification language for formally describing these services and an analysis technique for verifying correctness of the generated applications.

The paper by Greco, Leone, and Scarcello does not deal explicitly with networked data; the topic there is query optimization, specifically query decomposition based on hyper tree-width bounds. However, this kind of optimization is particularly useful in conjunction with XML queries (which often are guaranteed to have low tree width) and with relational queries generated from shredded XML documents. Another traditional database research topic that takes on new forms with networked access to data is security. The paper by Martens and Neven is clearly relevant to security in the broad sense, because it gives a method of gaining confidence that data exchange transformations will produce the output expected by data consumers. An important aspect of security, data confidentiality, is addressed in Miklau and Suciu’s article. They present a framework for analyzing whether confidential information (formalized as a set of queries) can be disclosed by a set of views.

Overall the papers in this issue provide an in-depth look at the challenges in making data available over networks. We thank all the authors of submitted papers, including those authors of papers that could not be included in this special issue due to reviewer revision requests that could not be accommodated in our tight time frame for publication. Special thanks are due to the editor-in-chief of *Journal of Computer and Systems Science*, Ed Blum, and the associated editor, Gorg Gottlob, for hosting this special issue.

Michael Benedikt
Bell Laboratories
263 Shuman Blvd.
Naperville, IL 60566, USA
E-mail address: benedikt@research.bell-labs.com

Available online 8 December 2006